An explicit formula for minimizing the infected peak in an SIR epidemic model when using a fixed number of complete lockdowns

Eduardo D. Sontag*

ABSTRACT

A too-early start of NPIs (non-pharmaceutical interventions) such as social distancing may lead to high "second waves" of infections of COVID-19. This paper asks what should be the timing of a set of k complete-lockdowns of prespecified lengths (such as two weeks) so as to minimize the peak of the infective compartment. Perhaps surprisingly, it is possible to give an explicit and easily computable rule for when each lockdown should commence. Simulations are used to show that the rule remains fairly accurate even if lockdowns are not perfect.

1 Introduction

The year 2020 will be remembered for the COVID-19 (coronavirus disease 2019) pandemic, which is an individual-to-individual infection by SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2). An immediate way to stop transmissibility is to establish total lockdowns, as China and Northern Italy did early on in the outbreak [1]. Lockdowns and other NPIs (non-pharmaceutical interventions) such as quarantine and social distancing were soon implemented around the world. On the other hand, frustration with isolation rules and economic costs mean that, in most countries, long lockdowns are not feasible, nor is it easy to enforce even milder forms of NPIs [2, 3, 4]. Moreover, a source of frustration in the COVID-19 epidemic was that a too-early start of NPIs often led to large "second waves" of infections when the NPIs where relaxed. This motivates the search for optimally timing the start of NPIs so as to minimize the maximum peak of infective individuals. This paper deals with that problem.

Infectious agents have critically influenced the history of mankind, with disease-causing pathogens constantly emerging or evolving. From the Plague of Athens (430-428 BC), to the fourteenth century Black Death that killed about a third of Europe's population, to the Yellow Fever epidemic in Philadelphia in 1793, in which a tenth of the population of the city perished, to the 1918 "Spanish flu" pandemic (which did not originate in Spain) that resulted in about 3-5% of the world population dying, to the COVID-19 2020-2021 pandemic, infectious diseases have had major impacts on health, psychological and social well-being, medical advances (mRNA vaccines, for example), economics, politics, military history, and religious and racial persecution. Different types of pathogens are involved in infectious diseases. Viruses cause the common cold, influenza, measles, West Nile, and COVID-19, while anthrax, salmonella, chlamydia ,and cholera are caused by bacteria, and protozoa give rise to malaria and trypanosomiasis (sleeping sickness). There are many mechanisms for transmission, including respiratory droplets (influenza, colds), body secretions (chlamydia), flies (trypanosomiasis), mosquitoes (malaria), and food or water (cholera). Control strategies include behavioral and sanitation changes (NPIs), vaccines, antibiotics, antiviral drugs. Notwithstanding this variety, there is a common mathematical structure.

The modeling of infectious diseases and their spread is an important part of mathematical biology, specifically mathematical epidemiology. Modeling is an important tool for gauging the impact of NPIs such as social distancing, masking, lock-downs, or school closings, as well as predicting/attenuating magnitude of peak infections ("flattening the curve" so as not to overwhelm ICU capacities), predicting/delaying peak

*Departments of Electrical and Computer Engineering and Bioengineering, and affiliate of Departments of Mathematics and Chemical Engineering, Northeastern University, Boston, MA, United States. Laboratory of Systems Pharmacology, Harvard Medical School, Boston, MA, United States. Email: sontag@sontaglab.org, ORCID: 0000-0001-8020-5783

infections (until vaccine/treatments available), and devising strategies for vaccination, control, or eradication of diseases. The social and political use of epidemic models must take into account their degree of realism. Good models do not incorporate all possible effects, but rather focus on the basic mechanisms in their simplest possible fashion. Not only it is difficult to model every detail, but the more details the more the likelihood of making the model sensitive to parameters and assumptions, and the more difficult it is to *understand and interpret* the model as well as to play "*what-if*" scenarios to compare alternative containment policies. It turns out that even simple models help pose important questions about the underlying mechanisms of infection spread and possible means of control of an epidemic.

Most mathematical epidemiology models incorporate some version of the classical SIR model proposed by Kermack and McKendrick in 1927 [5]. We will restrict attention to this core model, which is suitable for describing initial stages of an infection in a single city, and also for modeling later stages when community spread becomes dominant. Mathematical models have long played a central role in epidemiology, and this has been especially true with the COVID-19 pandemic [6, 7, 8, 9, 10, 11, 12, 13, 14]. This includes control-theoretic aspects, especially optimal control [15, 16, 17, 18].

Mathematical models had a major impact on the political response to the COVID-19 pandemic. To quote from "Behind the virus report that jarred the U.S. and the U.K. to action" (*New York Times*, 17 March 2020):

> *The report [from Imperial College London], which warned that an uncontrolled spread of the disease could cause as many as 510,000 deaths in Britain, triggered a sudden shift in the government's comparatively relaxed response to the virus. American officials said the report, which projected up to 2.2 million deaths in the United States from such a spread, also influenced the White House to strengthen its measures to isolate members of the public.*

Of course, one should always keep in mind as well the following quote from Dr. Anthony Fauci, Director, National Institute of Allergy and Infectious Diseases, United States (*CNN*, 05 May 2020):

> *I have skepticism about models [of COVID-19], and they are only as good as the assumptionsyou put into them, but they are not completely misleading. They are telling you something that is a reality, that when you have mitigation that is containing something, and unless it is down, in the right direction, and you pull back prematurely, you are going to get a rebound of cases.*

2 Formulation of problem

We consider a standard SIR model for epidemics, which consists of three groups of individuals: those who are *susceptible* and can be passed on the pathogen by the *infectious* individuals, and the *removed* individuals, who are have either developed immunity after infection or who have died. In SIR models, one does not include a flow back from individuals into the susceptible compartment. On longer timescales, one may also allow for the fact those individuals in the removed group who are immune may eventually return to the susceptible population, which would happen if immunity is only temporary or if a pathogen has evolved substantially. The numbers of individuals in the three classes will be denoted by S, I, and R respectively, and hence the name "SIR" model. When re-infection is possible (there is only temporary immunity), one talks about the "$SIRS$" model (the last S to indicate flow from R to S). Graphically, we draw these as follows, where we use the symbol "$\otimes$" to indicate that the number of new infected will depend both on the number is susceptibles and infected (specifically, it will be a product in the classical SIR model to be discussed next, with a proportionality constant β), and ν denotes the flow to

the recovered compartments. Observe that the "feedback" term implicit in the $\otimes$ effect means that this is not exactly a compartmental system, because for those, *the flow into* a compartment does not depend on the number of individuals in that compartment.

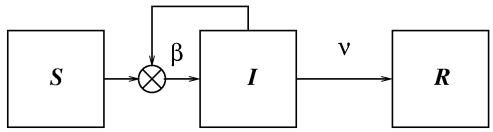

Figure 1: Diagram of standard SIR system

Mathematically, assuming that new infections are due to contacts between S and I individuals, and that the rate at which this happens is proportional to the numbers of such individuals, there results a system of three coupled ordinary differential equations as follows:

$$\dot{S} = -\beta SI \tag{1}$$

$$\dot{I} = \beta SI - \nu I \tag{2}$$

$$\dot{R} = \nu I \tag{3}$$

where from now on we will ignore the last equation since it does not affect the number of infectives. and initial state $S(0) = S_0 > 0$, $I(0) = I_0 > 0$. The positive parameter β (infectivity/contact rate) quantifies the transmission rate between the susceptible S and infected I individuals in a well-mixed population, and $!n$ (recovery/death rate) is the rate of flow into the removed R compartment. We will assume that $\mathcal{R}_0 := \beta S_0/\nu > 1$; otherwise, the problem to be discussed is not interesting, as I decreases monotonically to zero if the condition does not hold. (This and other well known facts about SIR systems are reviewed in Section 5; see also e.g. [19].)

In the SIR model, NPIs are viewed as reducing the contact rate β. (More sophisticated models of social distancing have been widely studied, see for example [20, 21, 22, 23, 24, 25, 26].) The reduction of β is modeled by a time-varying $\beta(t)$ where

$$\beta(t) = \begin{cases} \beta_0 & t \in J \\ \beta & t \notin J \end{cases} \tag{4}$$

where $0 \leq \beta_0 < \beta$ are two fixed values, and $J \subseteq [0, \infty)$. In our work, J will be the union of a number K of intervals

$$J_k = [t_k, t_k + T_k], \quad k \in K$$

and *we will be taking* $\beta_0 = 0$, representing a strict or full lockdown. The lengths T_k of the respective intervals will be allowed to be arbitrary (but fixed) in our theory, though the most elegant formulas are obtained when they are all equal (we might have $T_k = 14$ or 28 days for all k, for instance). Obviously, a perfect or near-perfect lockdown is not practical (except in certain countries where complete obedience is either culturally or militarily enforced), but this theoretical problem is nonetheless of interest. Moreover, it helps understand the problem with $\beta_0 > 0$, and we show via simulations that the conclusions for $\beta_0 = 0$ are very relevant to the case of a small but nonzero β_0, even one that is about 20% of the value of β. The objective will be to minimize the maximum of $I(t)$ for $t \geq 0$ ("flatten the curve") by appropriately choosing the start times t_k.

Previous work on this and related problems includes [27], which treats an optimal schedule minimizing a combination of the total number of deaths and the peak of the infected compartment, [17] which shows that a single interval ($K = 1$) is optimal if the objective is to minimize the total number of susceptible individuals at the end of the epidemic, [15] with numerical studies of optimally timing fixed-duration

"one-shot" strategies, and the very nice theoretical paper [16] which showed that the optimal strategy for minimizing peak infection is a combination of a strict lockdown ("full suppression") with a feedback strategy which keeps $\mathcal{R}_0$=1. Also closely related to this work is [28], which studies timing of lockdowns, including periodic strategies, through a combination of theoretical and numerical methods,

Let $r := \nu/\beta$, and consider the "virtual peak" of $I(t)$ if no lockdown were imposed:

$$V \;:=\; \max_{t\geq 0} I(t) \;=\; I_0 + S_0 - r(1 - \ln(r/S_0))$$

(cf. Section 5). Define this expression:

$$I_{\text{start}} \;:=\; \frac{V_0}{1 + k - (e^{-\tau T_1} + \ldots + e^{-\tau T_K})}$$

The main result is as follows:

Theorem. *Suppose that $I(0) < I_{start}$. Then, in order to minimize the maximum of $I(t)$, $t \geq 0$, lockdowns should start whenever*

$$I(t_k) \;=\; I_{\text{start}}\,.$$

Moreover, under this policy, the maximum of $I(t)$ will equal I_{start}.

In other words, any time that the infective population level reaches the value I_{start}, the next time-T_k lockdown interval I_k should start.

For example, if $T_k = T$ for $i = 1, \ldots, K$ then the formula is

$$I_{\text{max}} \;=\; \frac{V_0}{1 + k - ke^{-\tau T}}\,.$$

Note that as $T \to \infty$ the best possible maximum peak is

$$\frac{V_0}{1+k}\,.$$

We prove the theorem next.

For mathematical elegance, we will include the theoretical possibility of a lockdown starting at time exactly $t_0 = 0$ (and later make $T_0 = 0$, so that there is in effect no initial lockdown).

For any initial population (σ, ι), we introduce the following function:

$$\mathcal{V}(\sigma, \iota) \;:=\; \iota + \sigma - r + r\ln r - r\ln\sigma$$

which gives the peak value of $I(t)$ if we start at the initial population (σ, ι), and let:

$$\begin{aligned}
t_k &:= \text{time when } k\text{th lockdown starts}\\
I_k &:= I(t_k)\\
S_k &:= S(t_k)\\
V_k &:= \mathcal{V}(S_k, I_k)\\
T_k &:= \text{length of } k\text{th lockdown}\\
p_k &:= e^{-\nu T_k}
\end{aligned}$$

(so $t_0 = 0$ and (S_0, I_0) is the state at the start of the epidemic). Observe that

$$\begin{aligned}
I(t_k + T_k) &= p_k I_k\\
S(t_k + T_k) &= S_k\,.
\end{aligned}$$

Consider these equalities:

$$\begin{aligned} V_k + r - r\ln r &= p_k I_k + S_k - r\ln S_k && (5)\\ V_{k-1} + r - r\ln r &= p_{k-1} I_{k-1} + S_{k-1} - r\ln S_{k-1} && (6)\\ I_k + S_k - r\ln S_k &= p_{k-1} I_{k-1} + S_{k-1} - r\ln S_{k-1} && (7) \end{aligned}$$

where the last equality follows from the conservation law $I(t) + S(t) - r\ln S(t) \equiv$ constant, applied to a solution that starts at $\iota = p_k I_k$ and $\sigma = S_k$. Adding (5) to (7) and subtracting (6), we obtain:

$$V_k + I_k + S_k - r\ln S_k - V_{k-1} \;==\; p_k I_k + S_k - r\ln S_k$$

from which we conclude the following recursion:

$$V_k \;=\; -a_k I_k + V_{k-1}$$

where for convenience we let $a_k := 1 - p_k$. Applying this formula recursively,

$$V_K = V_0 - (a_0 I_0 + \ldots + a_K I_K)\,.$$

The largest value of $I(t)$ will occur at the maximum of the peaks occurring at the start of each kth lockdown, or the last "virtual peak" (which is then a real local maximum). In other words, the maximum of I is

$$I_{\text{max}} \;=\; \max\{I_0, I_1, \ldots, I_K, V_K\}\,.$$

Under the epidemic assumption $\mathcal{R}_0 > 1$, $\dot{I}(0) > 0$, so we can drop I_0 from the list. Similarly, we can replace the above value for V_K. letting $T_0 = 0$ means that $a_0 = 1 - p_0 = 0$, so we conclude that

$$I_{\text{max}} \;=\; \max\{I_1, \ldots, I_K, V_0 - (a_1 I_1 + \ldots + a_K I_K)\}\,.$$

To minimize the peak, we need to pick non-negative $I_1, \ldots, I_K$ in such a way that this expression is minimized. This is achieved at a unique local minimum, at which

$$I_{\text{max}} = I_1 = \ldots = I_K = V_0 - (a_1 I_1 + \ldots + a_K I_K) \qquad \text{for all } k$$

(see subsection 2.1 below), namely

$$I_k \;=\; \frac{V_0}{1 + (a_1 + \ldots + a_K)} \;=\; \frac{V_0}{1 + k - (e^{-\tau T_1} + \ldots + e^{-\tau T_K})} \qquad \text{for all } k\,.$$

This completes the proof of the theorem. ∎

2.1 A small lemma

Let $a_1 \geq 0, \ldots, a_K \geq 0$, $b > 0$. Consider the following function, defined on $(x_1, \ldots, x_K) \in \mathbb{R}^K_{\geq 0}$:

$$g(x_1, \ldots, x_k) \;:=\; \max\left\{x_1, \ldots, x_K, b - \sum_{k=1}^{K} a_k x_k\right\}\,.$$

Let $A := \sum_{k=1}^K a_k$ and $\xi := \frac{b}{1+A}$. Note these properties:

(1) When all $x_k = \xi$, $g(x_1, \ldots, x_k) = \xi$. To prove this, note that $b - \sum_{k=1}^K a_k x_k = b - A\xi = b - A\frac{b}{1+A} = \xi$, so all terms in the max are the same.

(2) If $g(x_1, \ldots, x_k) \leq \xi$ then (by definition of the max) (a) $x_k \leq \xi$ for all k, and (b) $b - \sum_{k=1}^K a_k x_k \leq \xi$.

(3) If $g(x_1, \ldots, x_k) < \xi$ for some $(x_1, \ldots, x_k)$, then at least one of the inequalities (a) or (b) must be is strict. Note that (a) implies that

$$b - \sum_{k=1}^{K} a_k x_k \geq b - A\xi = b - A\frac{b}{1+A} = \xi$$

and this is a strict inequality if one of the inequalities in (a) is strict, which then contradicts (b). If instead (b) is strict, then $b - \sum_{k=1}^{K} a_k x_k \geq \xi$ is contradicted. In summary, $g(x_1, \ldots, x_k) \geq \xi$ for all $(x_1, \ldots, x_k)$.

We conclude from (1) and (3) that ξ is the minimum value of g, and it is achieved when all $x_k = \xi$. We prove next that the minimum is achieved only at this point. Indeed, suppose that $g(x_1, \ldots, x_k) = \xi$, so that (a) and (b) above hold. If any $x_k \neq \xi$, then $x_k < \xi$, because if $x_k > \xi$ for any k, then $g(x_1, \ldots, x_k) > \xi$ (by definition of g). But we already remarked that this contradicts (b).

3 Optimal reduction of "virtual peak" with k perfect lockdowns

The use of a small number of lockdowns results in a drastic reduction of the peak that would occur (the "virtual pear") if there were no lockdowns. The marginal benefit of additional lockdowns is relatively minor, after a certain number of them. We plot here the fraction $\left(1 + k(1 - e^{-\nu T})\right)^{-1}$ for $\nu = 0.05$ and $T = 7, 14, 21, 28$ days.

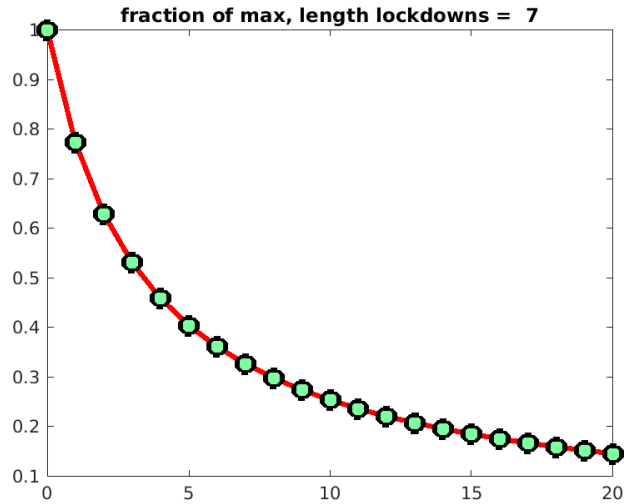

Figure 2: Percentage of virtual peak when using k lockdowns (horizontal axis) and $T = 7$ days

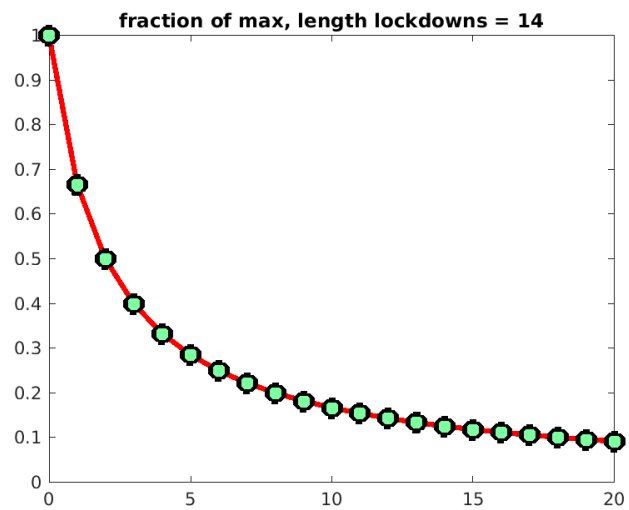

Figure 3: Percentage of virtual peak when using k lockdowns (horizontal axis) and $T = 14$ days

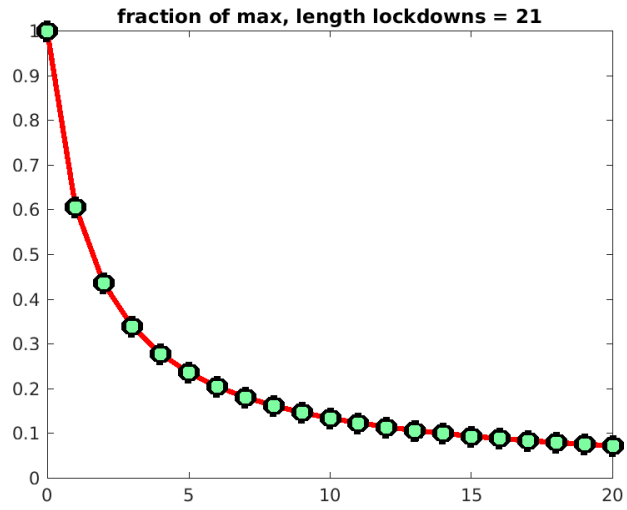

Figure 4: Percentage of virtual peak when using k lockdowns (horizontal axis) and $T = 21$ days

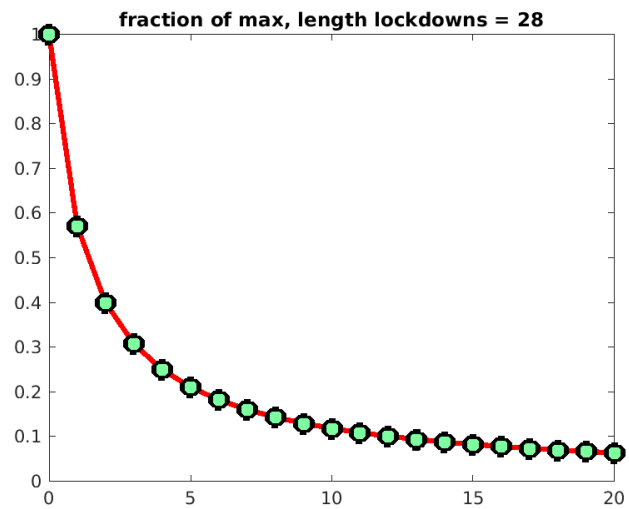

Figure 5: Percentage of virtual peak when using k lockdowns (horizontal axis) and $T = 28$ days

4 Simulations

4.1 Simulations using optimal formula

We simulate various lockdown lengths using our optimal formulas. Parameters are, $\beta = 0.00025$, $\nu = 0.05$ (so $\mathcal{R}_0 = 5$), initial conditions are $S_0 = 1000$, $I_0 = 1$, and the peak if there are no lockdowns is $I = 479$.

4.1.1 Lockdown length is $T = 14$ days, perfect lockdown

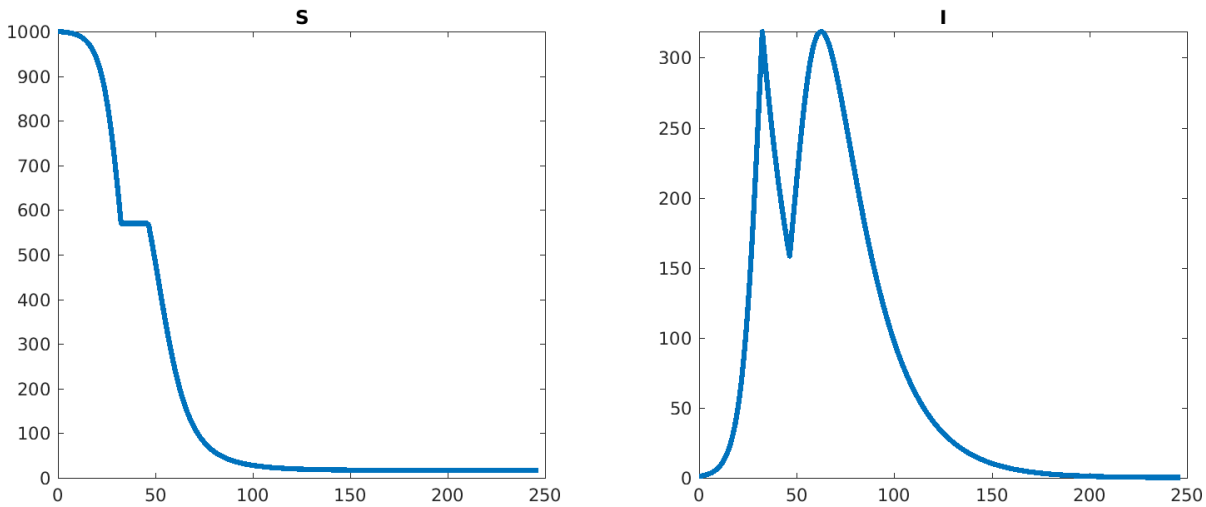

Figure 6: # lockdowns: 1, peak I from formula: 318.682808, maximum of I on last (no lockdown) period: 318.601530. Lockdown start time(s): 32.42.

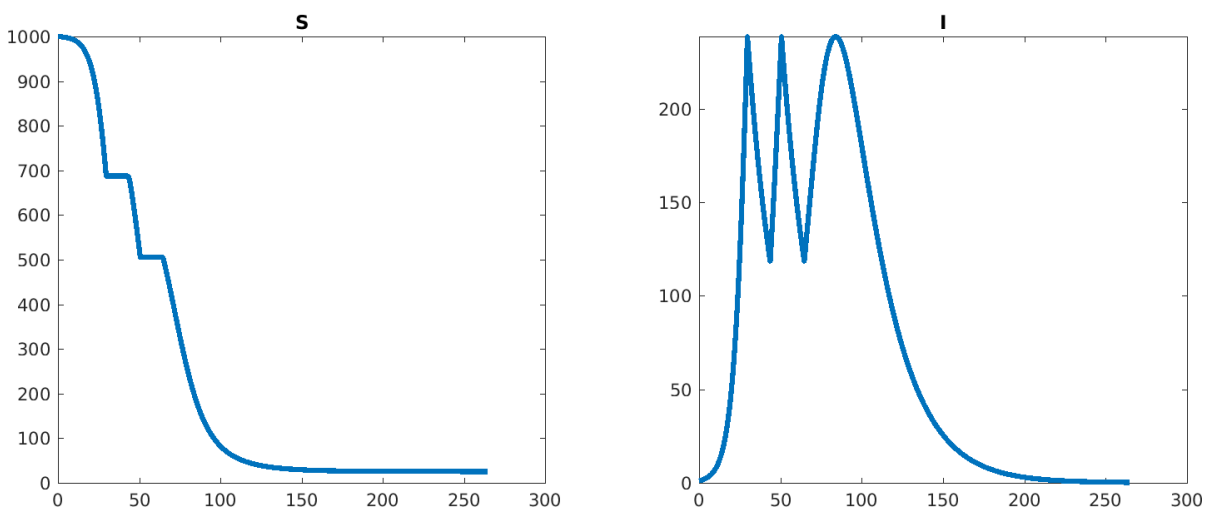

Figure 7: # lockdowns: 2, peak I from formula: 238.740981, maximum of I on last (no lockdown) period: 238.755779. Lockdown start time(s): 29.73, 50.69.

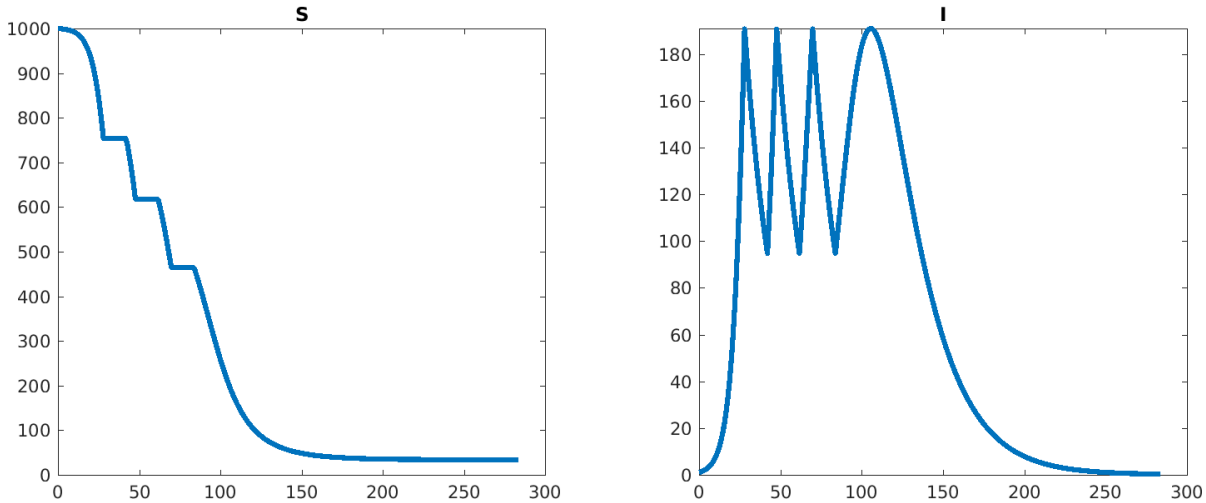

Figure 8: # lockdowns: 3, peak I from formula: 190.862880, maximum of I on last (no lockdown) period: 191.019085. Lockdown start time(s): 28.01, 47.71, 69.80.

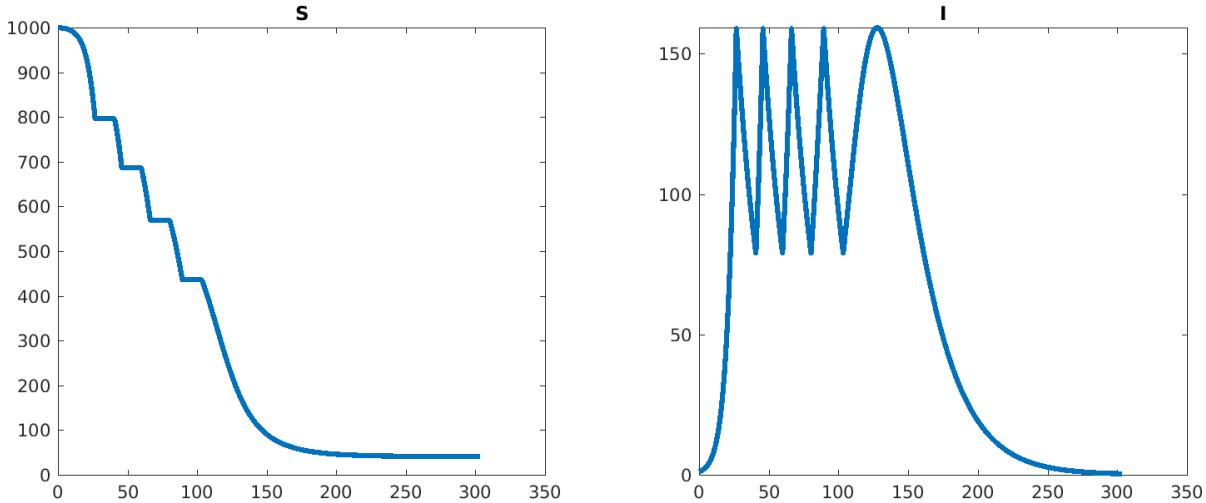

Figure 9: # lockdowns: 4, peak I from formula: 158.980313, maximum of I on last (no lockdown) period: 159.342577. Lockdown start time(s): 26.74, 45.87, 66.33, 89.44.

4.1.2 Lockdown length is $T = 28$ days, perfect lockdown

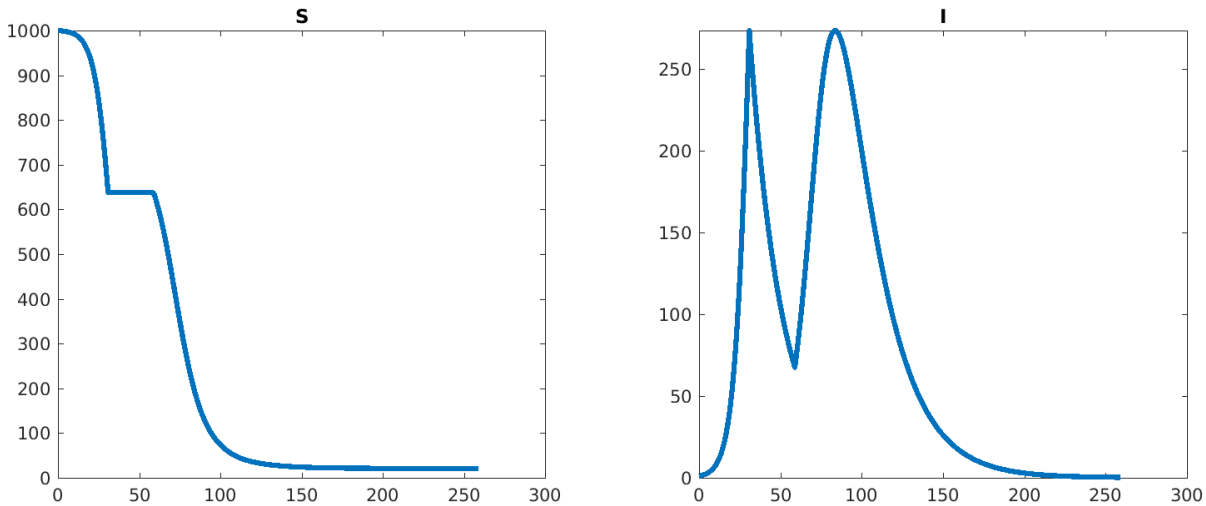

Figure 10: # lockdowns: 1, peak I from formula: 273.247170, maximum of I on last (no lockdown) period: 273.286548. Lockdown start time(s): 30.9.

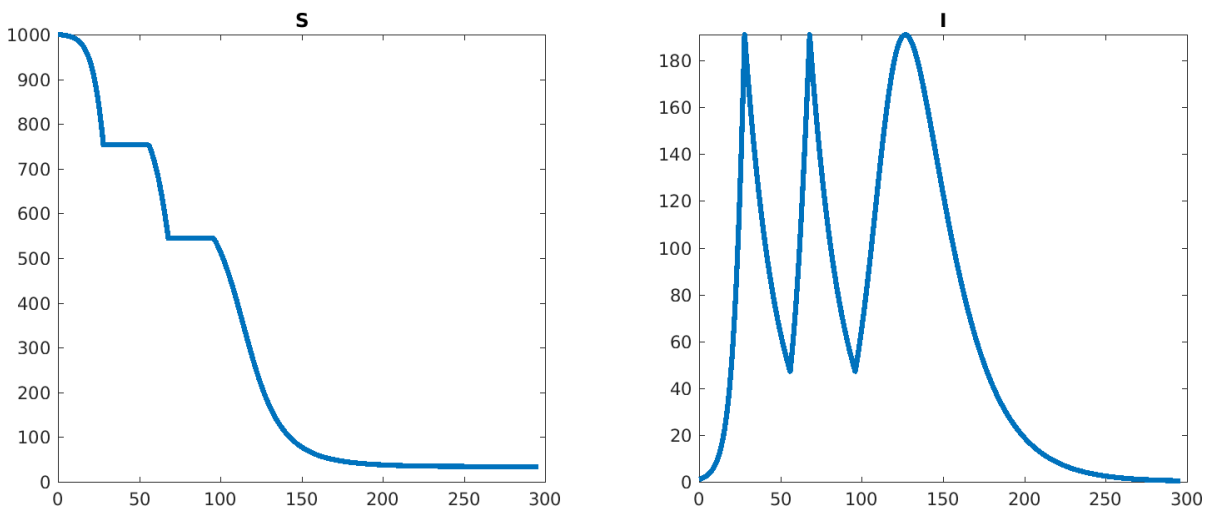

Figure 11: # lockdowns: 2, peak I from formula: 191.124644, maximum of I on last (no lockdown) period: 191.048654. Lockdown start time(s): 28.02, 68.02.

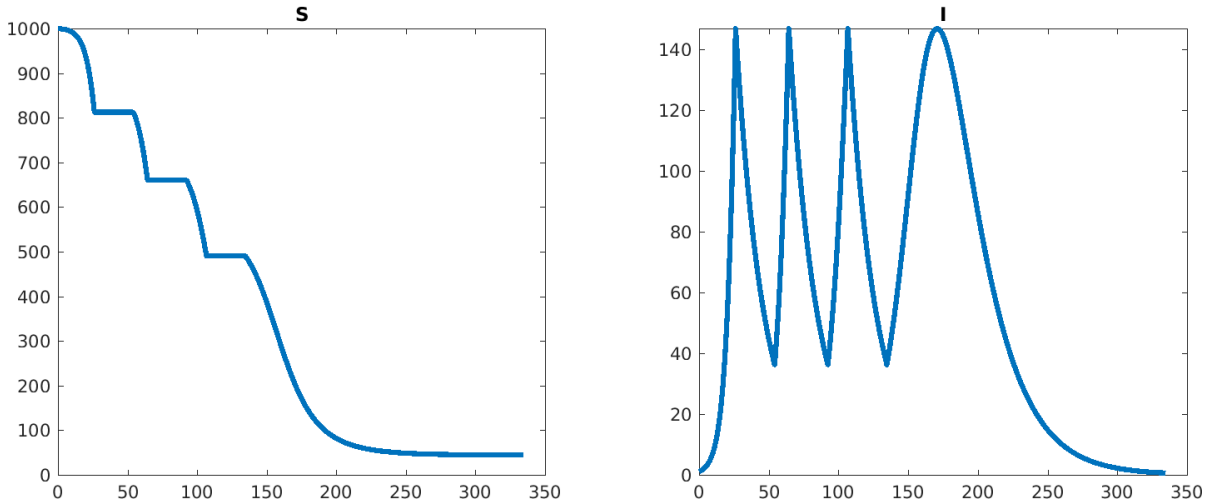

Figure 12: # lockdowns: 3, peak I from formula: 146.957573, maximum of I on last (no lockdown) period: 146.913773. Lockdown start time(s): 26.22, 64.39, 106.74.

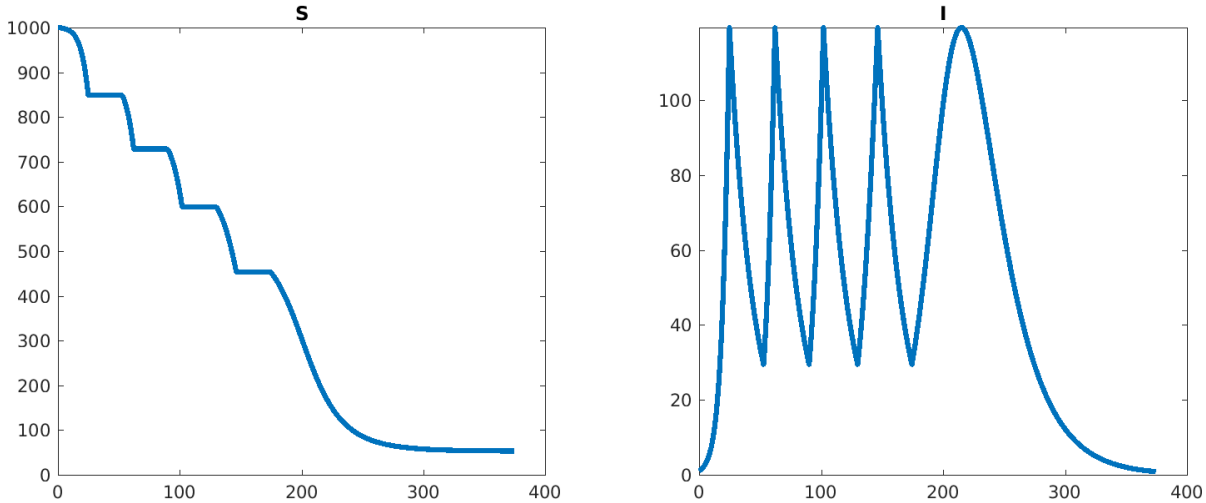

Figure 13: # lockdowns: 4, peak I from formula: 119.371878, maximum of I on last (no lockdown) period: 119.395364. Lockdown start time(s): 24.91, 62.23, 101.98, 146.47.

4.2 Testing $\beta_0 \neq 0$ with formula derived for $\beta_0 = 0$

We simulate the use of the formula derived for $\beta_0 = 0$, under the lockdown value $\beta_0 = 0.00005$, which represents a 20% value of the normal contact rate. Again, $\beta = 0.00025$, $\nu = 0.05$ (so $\mathcal{R}_0 = 5$), initial conditions are $S_0 = 1000$, $I_0 = 1$, and the peak if there are no lockdowns is $I = 479$.

4.2.1 Lockdown length is $T = 28$ days, $\beta_0 = 0.00005$

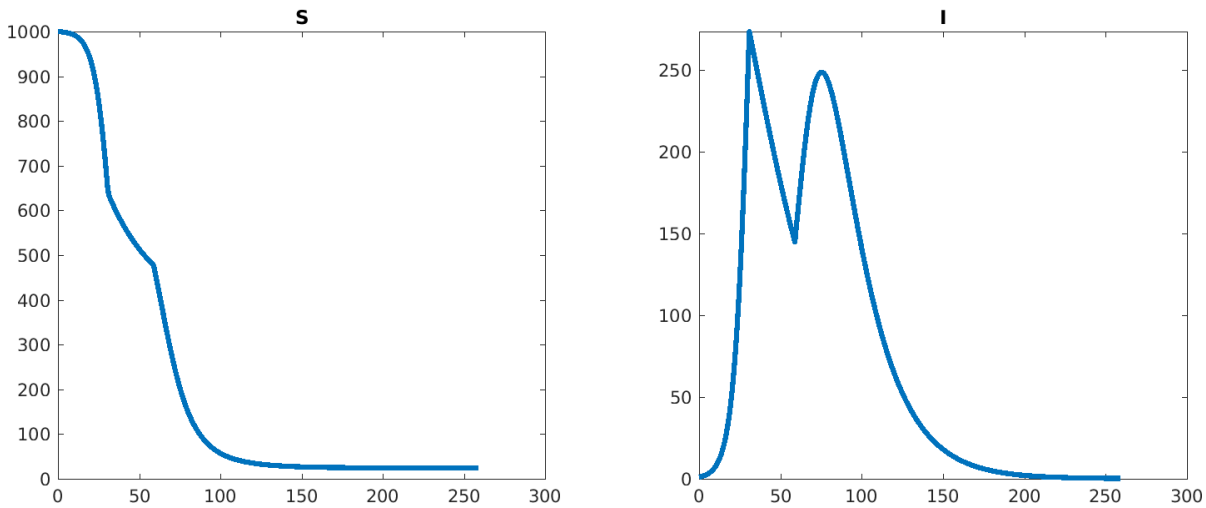

Figure 14: # lockdowns: 1, peak I from formula: 273.247170, maximum of I on last (no lockdown) period: 248.383407. Lockdown start time(s): 30.90.

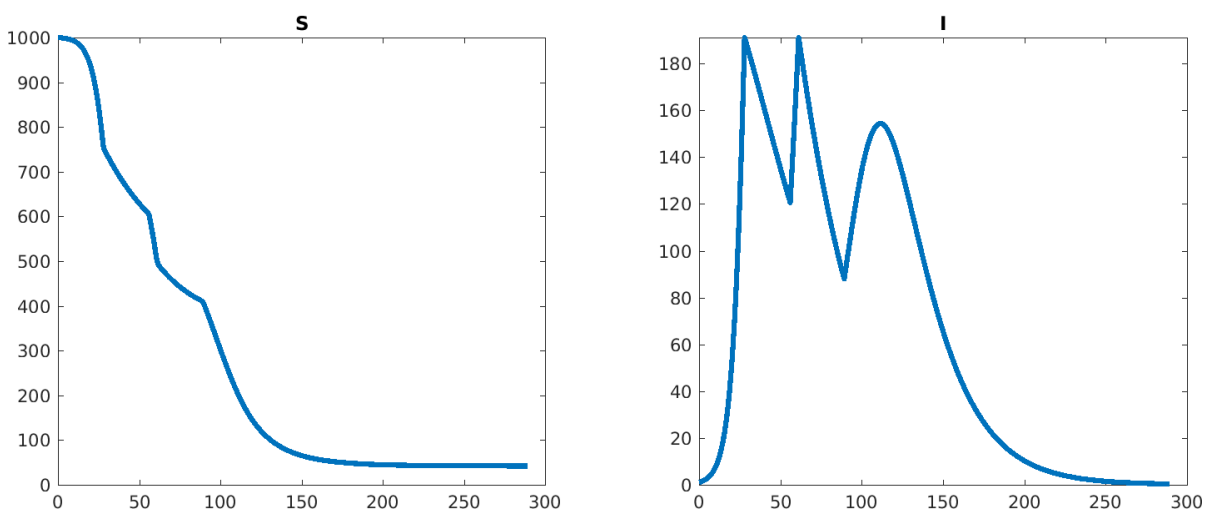

Figure 15: # lockdowns: 2, peak I from formula: 191.124644, maximum of I on last (no lockdown) period: 154.387915. Lockdown start time(s): 28.02, 61.2.

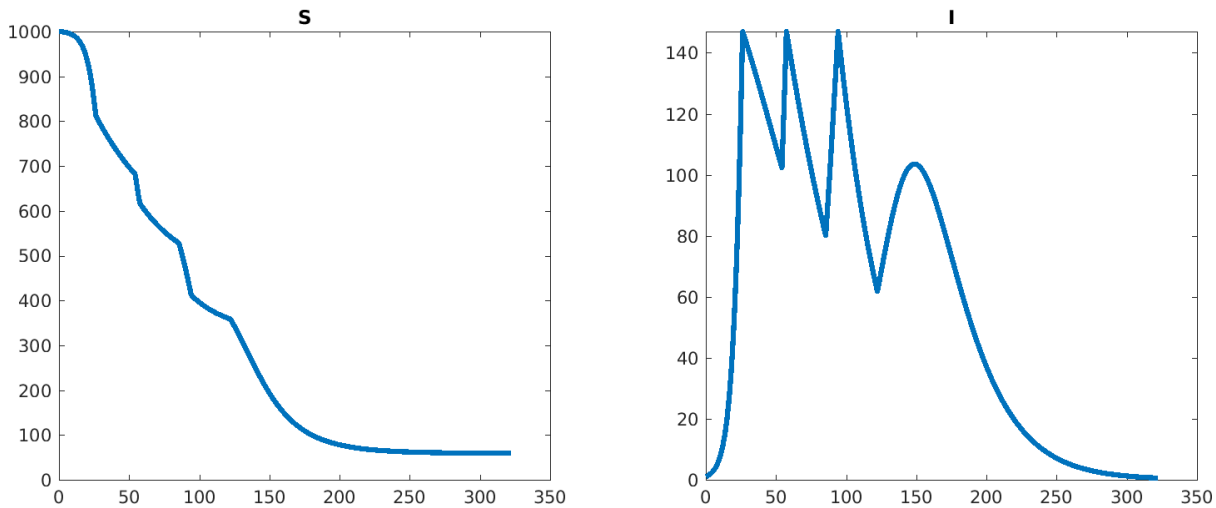

Figure 16: # lockdowns: 3, peak I from formula: 146.957573, maximum of I on last (no lockdown) period: 103.506053. Lockdown start time(s): 26.22, 57.43, 94.30.

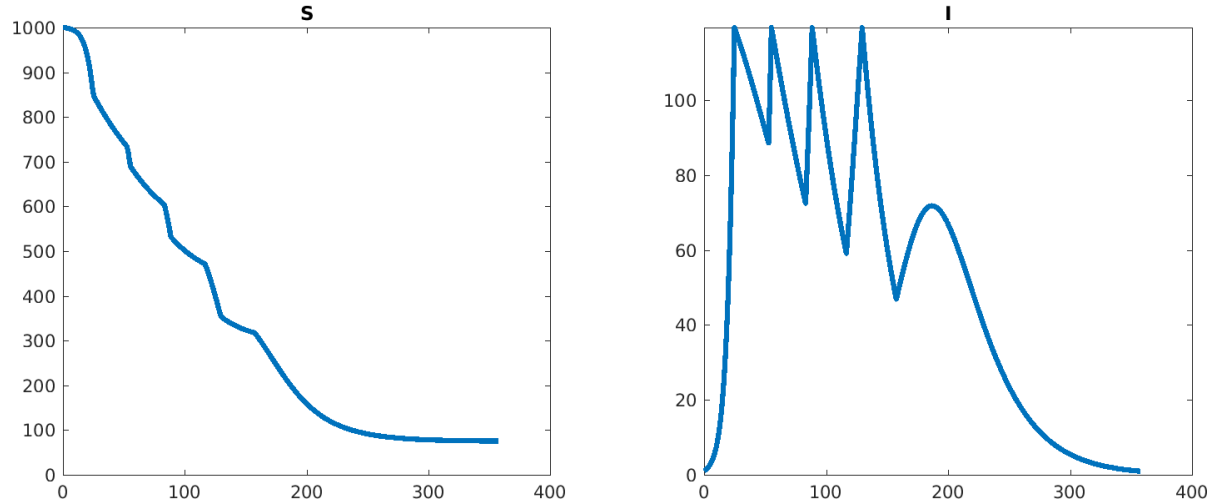

Figure 17: # lockdowns: 4, peak I from formula: 119.371878, maximum of I on last (no lockdown) period: 71.792718. Lockdown start time(s): 24.91, 55.24, 88.63, 129.63.

4.2.2 Lockdown length is $T = 14$ days, $\beta_0 = 0.00005$

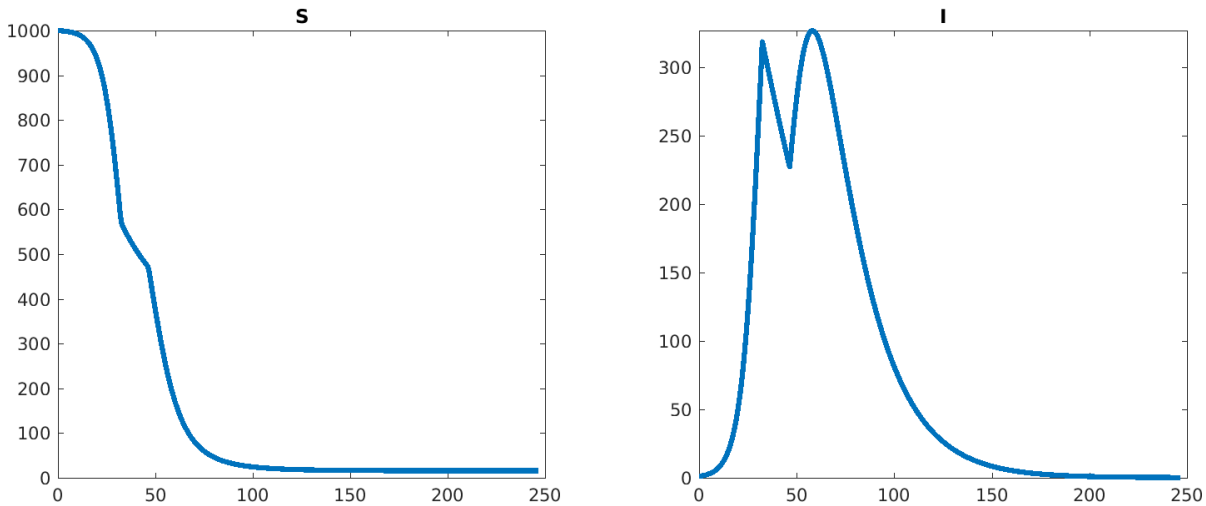

Figure 18: # lockdowns: 1, peak I from formula: 318.682808, maximum of I on last (no lockdown) period: 326.846639. Lockdown start time(s): 32.4.

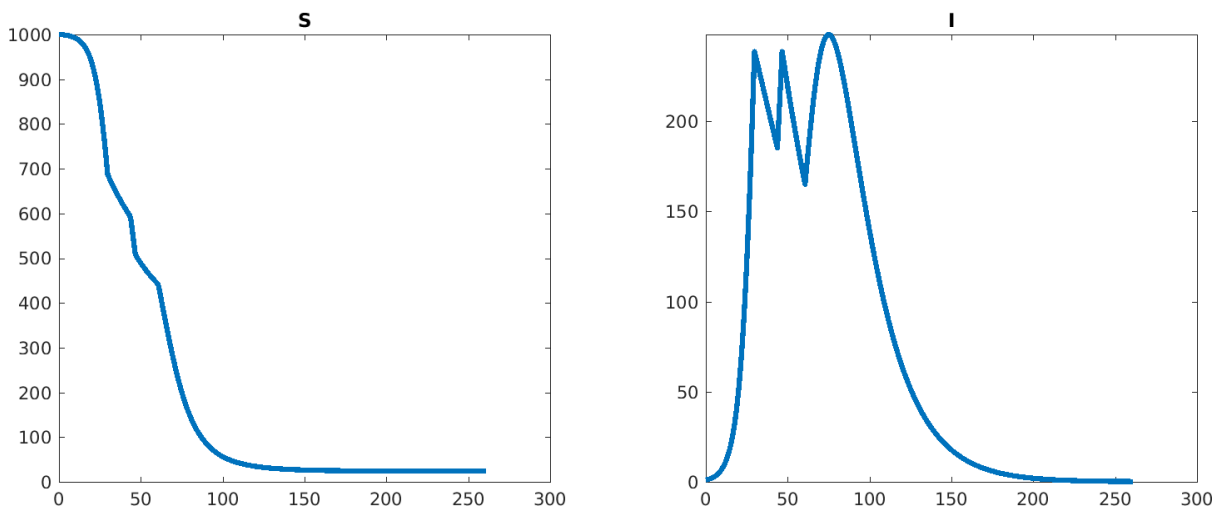

Figure 19: # lockdowns: 2, peak I from formula: 238.740981, maximum of I on last (no lockdown) period: 248.153424. Lockdown start time(s): 29.73, 46.6.

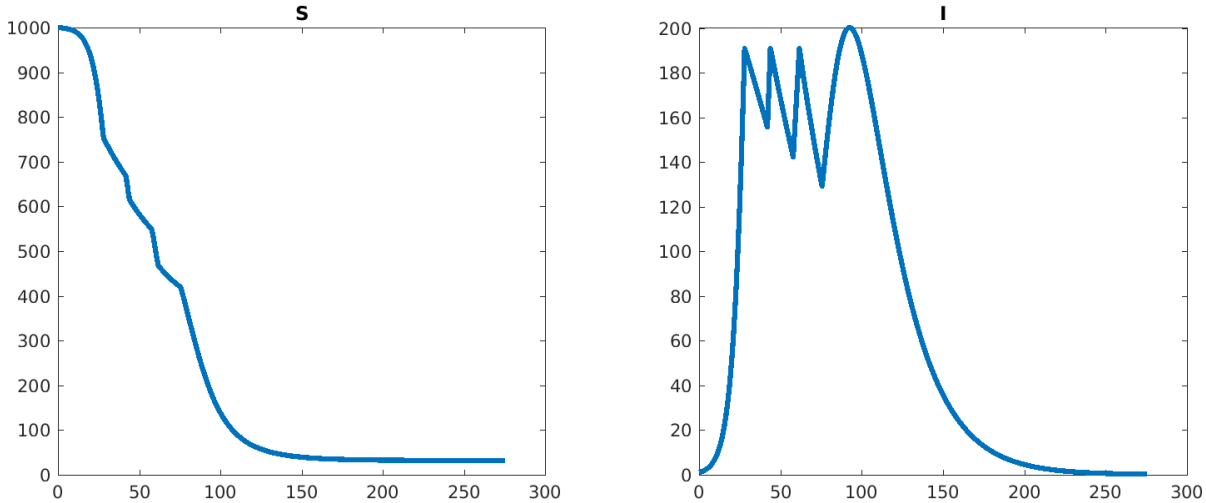

Figure 20: # lockdowns: 3, peak I from formula: 190.862880, maximum of I on last (no lockdown) period: 200.218534. Lockdown start time(s): 28.01, 43.86, 61.6.

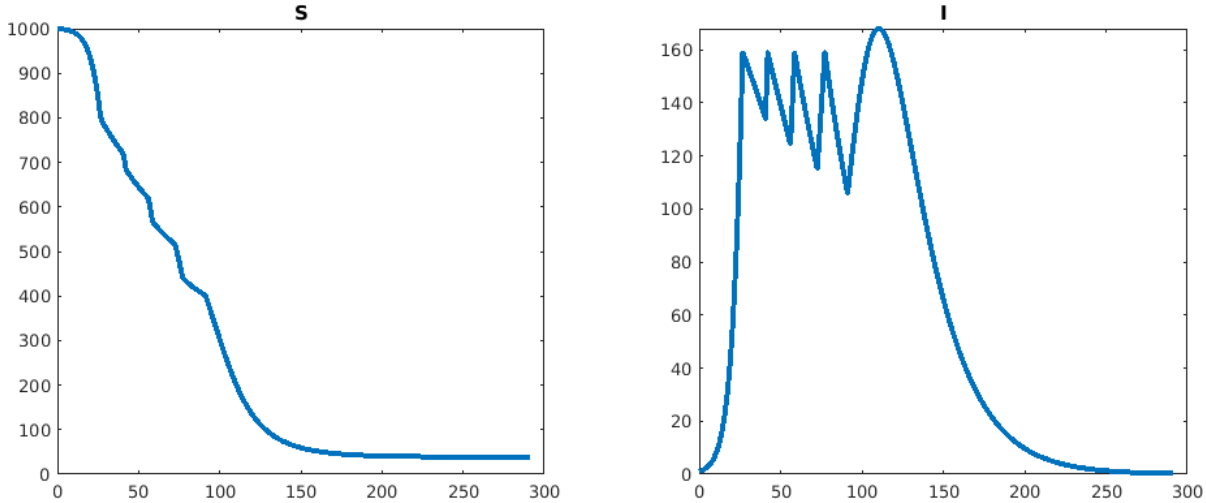

Figure 21: # lockdowns: 4, peak I from formula: 158.980313, maximum of I on last (no lockdown) period: 167.977200. Lockdown start time(s): 26.74, 42.11, 58.60, 77.22.

4.3 Comparison with optimal strategy for $\beta_0 \neq 0$

For comparison with the use of the optimal formulas derived for $\beta_0 = 0$, we show here numerically the optimal solution when $\beta_0 \neq 0$, specifically $\beta_0 = 0.00005$ as above. As earlier, parameters are $\beta = 0.00025$, $\nu = 0.05$ (so $\mathcal{R}_0 = 5$), initial conditions are $S_0 = 1000$, $I_0 = 1$, and the peak if there are no lockdowns is $I = 479$. We take only the case of a single lockdown, for simplicity, and lockdown lengths of 14 or 28 days.

We find that the formula predicts the optimal timing extremely well for 14-day lockdowns (error less than 1% in maximum infectives), and is fairly good for 28-day lockdowns as well (about 5% error).

4.3.1 Lockdown length is $T = 14$ days, $\beta_0 = 0.00005$

For 14-day lockdowns, the plot in Figure 18 suggests that the formula derived for $\beta_0 = 0$ triggered the first lockdown too early, so we explored by what fraction > 1 to increase the trigger point, from which the optimal strategy is clear.

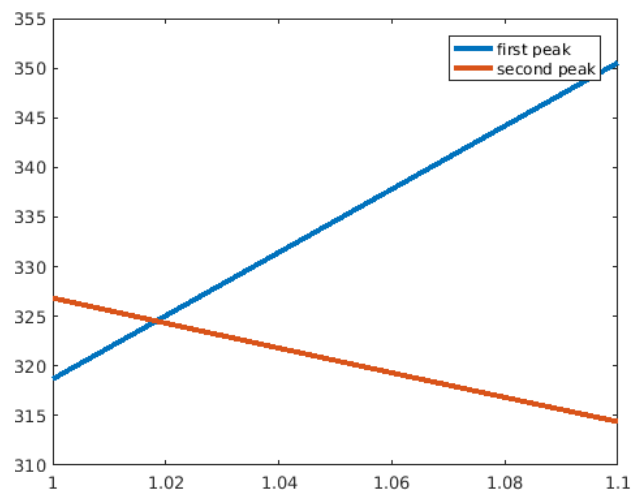

Figure 22: Magnitude of first and second peak, with the lockdown time parametrized by the percentage of the ideal (perfect lockdown) formula. Note that the second peak decreases as the first peak happens later, so the minimum of the maximum among them will occur when the curves intersect.

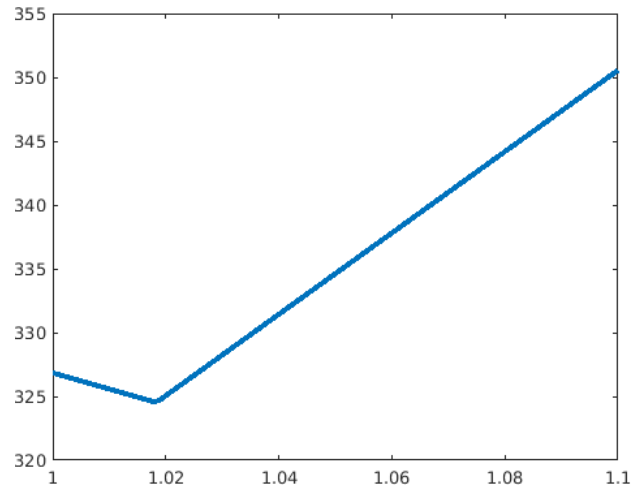

Figure 23: The maximum between the curves in Figure 26.

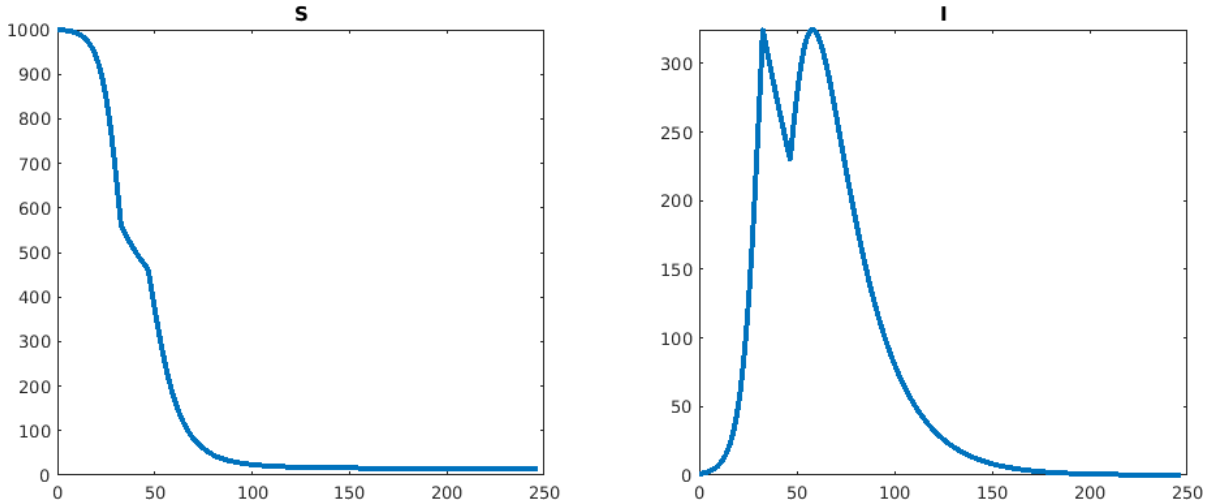

Figure 24: Simulation with an imperfect lockdown. Using now the optimal fraction 1.018 of the optimal for perfect lockdowns, obtained by minimizing the plot in Figure 23. The optimal peak value is now approximately 324, which can be compared with the suboptimal plot (using the formula that assumed perfect lockdowns) shown in Figure 18, which had a peak of approximately 327. Observe how the two peaks are now balanced. The optimal result is not that different from the one obtained from our formulas, and the lockdown start is at time 30.39.

4.3.2 Lockdown length is $T = 14$ days, $\beta_0 = 0.00005$

In contrast, for 28-day lockdowns the plot in Figure 14 suggests that the formula derived for $\beta_0 = 0$ waited too long for the first lockdown, so we first explore by what fraction < 1 to decrease the trigger point, from which the optimal strategy is clear.

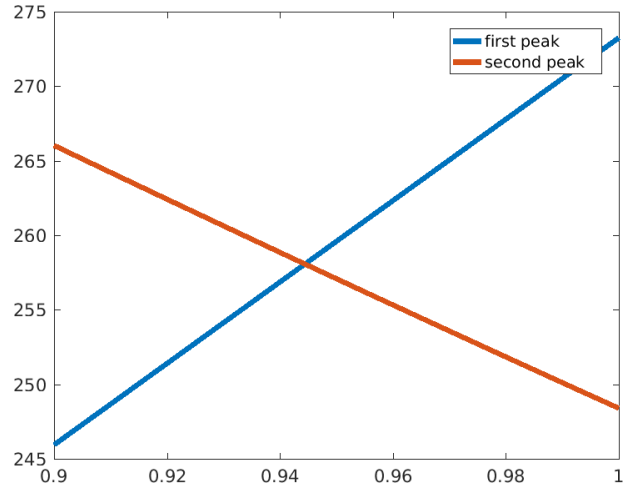

Figure 25: Magnitude of first and second peak, with the lockdown time parametrized by the percentage of the ideal (perfect lockdown) formula. Note that the second peak decreases as the first peak happens later, so the minimum of the maximum among them will occur when the curves intersect.

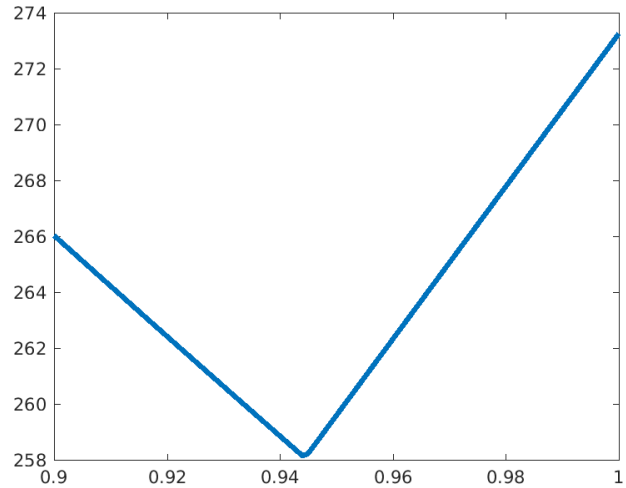

Figure 26: The maximum between the curves in Figure 26.

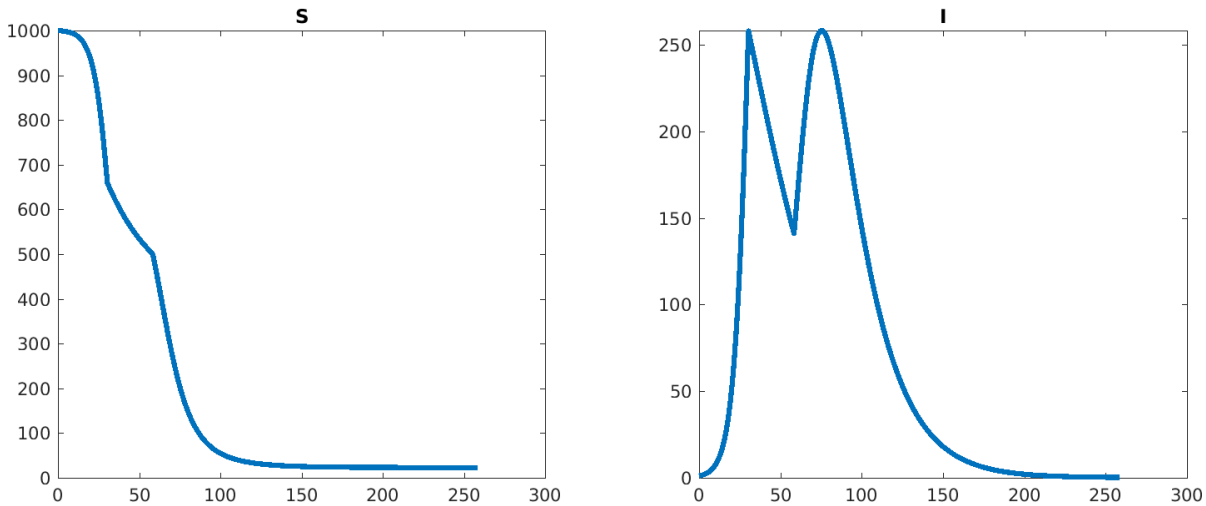

Figure 27: Simulation with an imperfect lockdown. Using now the optimal fraction 0.944 of the optimal for perfect lockdowns, obtained by minimizing the plot in Figure 26. The optimal peak value is now approximately 258, which can be compared with the suboptimal plot (using the formula that assumed perfect lockdowns) shown in Figure 14, which had a peak of approximately 273. Observe how the two peaks are now balanced. The optimal result is very close to the one obtained from our formulas.

5 Review of SIR model

We review here some facts about SIR models, see e.g. [19] for more details. We wish to analyze solutions, from initial conditions $S(0) = S_0$, $I(0) = I_0$.

Infections always die-out in SIR model

Note that if $I_0 = 0$, then $S(t) \equiv S_0$ and $I(t) \equiv 0$; in other words, every point of the form $(S, 0)$ is an equilibrium. Similarly, if $S_0 = 0$, then $S(t) \equiv 0$ and $I(t) = e^{-\nu t} I_0 \to 0$, so the case $S_0 = 0$ is not interesting either. So we study the only interesting cases, $I_0 > 0$ and $S_0 > 0$.

Since $\dot{S} \leq 0$, $S(t)$ is a nonincreasing function of time, and thus $S(t) \searrow S_\infty$ for some $S_\infty \geq 0$. A most important result is this one:

$$S_\infty > 0 \ \text{ and } \ I(t) \to 0 \ \text{ as } \ t \to \infty.$$

This says that the infection will end (asymptotically), and there will remain a number of "naive" individuals at the end.

We will show that $I(t) \to 0$ and defer the proof that $S_\infty > 0$ to later.

To prove this result, we will use this theorem: *if $x(t)$ is a solution of a system of ODEs $\dot{x} = f(x)$, and if the solution converges, $x(t) \to x^*$, then x^* must be an equilibrium point, i.e. $f(x^*) = 0$.* This is true because the omega-limit set of a trajectory is an invariant set, the LaSalle Invariance Principle (see e.g. [29]).

We apply this theorem as follows. First we define $V(t) := S(t) + I(t)$, and notice that $\dot{V} = -\nu I \leq 0$, which means that $V(t)$ is nonincreasing, and thus there is a limit $V(t) \searrow V_\infty$ as $t \to \infty$. Therefore $I(t) = V(t) - S(t) \to V_\infty - S_\infty =: I_\infty$ also has a limit. So the state $x(t) = (S(t), I(t))$ converges to $x^* := (S_\infty, I_\infty)$. It follows that $f(x^*) = 0$, which means that

$$\begin{aligned} \beta S_\infty I_\infty &= 0 \\ \beta S_\infty I_\infty - \nu I_\infty &= 0 \end{aligned}$$

and from there we conclude that $I_\infty = 0$ because $\nu \neq 0$.

We still have to show that $S_\infty > 0$; we will in fact provide a formula for S_∞ .

$\mathcal{R}_0$ and epidemics

A central role in epidemiology is played by the "intrinsic reproductive rate"

$$\mathcal{R}_0 := \frac{\beta S_0}{\nu} .$$

The epidemiological (and non-mathematically rigorous) definition of $\mathcal{R}_0$ is "the average number of secondary cases produced by one infected individual introduced into a population of susceptible individuals," where by a susceptible individual one means one who can acquire the disease. This can be made precise with a stochastic model, but an intuitive argument can be found in [19]. We remark $\mathcal{R}_0$ has a generalization to more complex epidemics models, and it characterizes the local stability of the set of "disease-free" steady states (DFSS) (those for there are no infectives). One may compute $\mathcal{R}_0$ using the so-called "next generation matrix" built from the differential equations. which was introduced in [30] (see e.g [31] and also the worked examples in [19]).

We discuss $\mathcal{R}_0$ below in more detail, but for now note the following fact. From the ODE for I, we have that

$$\frac{dI}{dt}(0) \;=\; (\beta S_0 - \nu) I_0 \;=\; \nu\,(\mathcal{R}_0 - 1) I_0\,.$$

This means that *an epidemic will happen*, meaning that $I(t)$ will increase when starting from any $I(0) > 0$, *if and only if* $\mathcal{R}_0 > 1$.

Moreover, the initial growth of $I(t)$ will be exponential, with rate $r = \nu\,(\mathcal{R}_0 - 1)$. (For small times and a large susceptible population S_0, we may assume that $S(t)$ remains roughly constant.) Logarithmically plotting infections, we can estimate r, and from there we may estimate

$$\mathcal{R}_0 \;=\; 1 + \frac{r}{\nu}$$

(assuming that we know ν, the recovery/death rate of infecteds), and

$$\beta \;=\; \frac{r+\nu}{S_0}\,.$$

When $\mathcal{R}_0 \leq 1$ and $t > 0$,

$$\frac{dI}{dt}(t) \;=\; (\beta S(t) - \nu) I(t) \;<\; (\beta S_0 - \nu) I \;=\; \nu\,(\mathcal{R}_0 - 1) I \;\leq 0$$

(because $S(t) < S_0$) and so $I(t)$ monotonically decreases to zero.

From now on, when discussing the SIR model, we assume that $\mathcal{R}_0 > 1$.

Peak infection time t_p and susceptibles at that time

The derivative $\dot{I} = (\beta S - \nu) I)$ is positive for small t, because at zero it equals $\nu\,(\mathcal{R}_0 - 1) I > 0$.

On the other hand, since $I(t) \to 0$ as $t \to \infty$, the derivative must eventually become negative, which means that there is some time t_p (p for "peak" infectivity) at which $\dot{I}(t_p) = 0$. Since $S(t)$ decreases monotonically, the derivative of I can only change sign from positive to negative at exactly one such time t_p. So t_p is the point at which $I(t)$ attains is maximum.

From $\dot{I} = 0$ at t_p, we have that

$$S(t_p) \;=\; \frac{\nu}{\beta}$$

or equivalently, dividing by the initial population,

$$\frac{S(t_p)}{S_0} \;=\; \frac{\nu}{\beta S_0} \;=\; \frac{1}{\mathcal{R}_0}$$

which says that at the peak infection time, we have a precise formula for the number of susceptibles. (Later, we give a value for $I(t_p)$ as well.)

A formula for the final number $S_\infty > 0$ of susceptibles

Let us know derive an (implicit) equation for the limit S_∞ of the susceptible population. We introduce the following function, along a given solution:

$$H(t) := \; I(t) + S(t) - \frac{\nu}{\beta} \ln S(t)\,.$$

Taking derivatives,

$$\frac{dH}{dt} \;=\; \beta SI - \nu I - \beta SI - \frac{\nu}{\beta}\frac{(-\beta SI)}{S} \;=\; 0$$

which means that H is constant along trajectories (a conserved quantity):

$$I(t) + S(t) - \frac{\nu}{\beta}\ln S(t) \;=\; I_0 + S_0 - \frac{\nu}{\beta}\ln S_0$$

for all $t > 0$. It follows, in particular, that

$$\frac{\nu}{\beta}\ln S(t) \;=\; I(t) + S(t) - I_0 - S_0 + \frac{\nu}{\beta}\ln S_0 \;\geq\; -I_0 - S_0 + \frac{\nu}{\beta}\ln S_0 \;=:\; p$$

and therefore $S(t) \geq e^{p\beta/\nu}$ for all t, so taking limits $S_\infty \geq e^{p\beta/\nu} > 0$ as claimed.

Even better, we can obtain an equation for S_∞ by passing to the limit in the conservation law, which gives (taking into account that $I_\infty = 0$):

$$S_\infty - \frac{\nu}{\beta}\ln S_\infty + \frac{\nu}{\beta}\ln S_0 \;=\; I_0 + S_0\,.$$

Dividing by S_0 and using that $\mathcal{R}_0 = \beta S_0/\nu$, we obtain:

$$\boxed{\frac{S_\infty}{S_0} - \frac{1}{\mathcal{R}_0}\ln\left(\frac{S_\infty}{S_0}\right) \;=\; 1 + \frac{I_0}{S_0}}$$

or, letting $x := S_\infty/S_0$ and $c := 1 + \frac{I_0}{S_0} > 1$:

$$f(x) := \; x - \frac{1}{\mathcal{R}_0}\ln x \;=\; c\,.$$

Observe that, since $S(t)$ is decreasing, $x < 1$. We claim that there is exactly one solution of the equation $f(x) = c$ with $x \in (0,1)$. By computing this solution, we can retrieve the final value of the susceptibles, $S_\infty = xS_0$. To prove that there is a solution x and it is unique, note that $\lim_{x\to 0^+} f(x) = +\infty$ and $f(1) = 1$, and $f'(x) = 1 - \frac{1}{\mathcal{R}_0}\frac{1}{x}$ is an increasing function of x, with $\lim_{x\to 0^+} f'(x) = -\infty$ and $f'(1) = 1 - \frac{1}{\mathcal{R}_0} > 0$ (because we assumed $\mathcal{R}_0 > 1$). Therefore, f decreases until some x^* and then increases back to 1. Since $c > 1$, it follows that $f(x) = c$ has a unique solution, as we wanted to prove.

There is in fact a solution of this equation that employs a classical function. For simplicity let us write $r = \mathcal{R}_0$. Multiplying by $-r$, we write the equation as $\ln x - rx = -rc$. Taking exponentials and multiplying again by $-r$ results in $we^w = y$, where $w := -rx$ and $y := -re^{-rc}$. Note that, since $r > 1$ and $c > 1$, $y \in (-1/e, 0)$. The function $w \mapsto we^w$ has an inverse, defined on $(-1/e, 0)$, called the *Lambert W function* (MATLAB command `lambertw`). So, $w = W(y)$, and since $x = -w/r$, we conclude that

$$x \;=\; -\frac{1}{r}W\left(-re^{-rc}\right)$$

where $r = \mathcal{R}_0$.

Typically, $I_0 \approx 0$ (one individual is enough to cause an epidemic), so we have the almost-exact formula

$$\frac{S_\infty}{S_0} - \frac{1}{\mathcal{R}_0}\ln\left(\frac{S_\infty}{S_0}\right) \;=\; 1\,.$$

If we measure the proportion of people who did not get sick compared to the total initial population, we can solve this equation for $\mathcal{R}_0$. This is one way that people compute $\mathcal{R}_0$ for historical diseases.

A formula for the peak value $I(t_p)$ of infectives

Determining the peak value $I(t_p)$ is of critical importance in practice. If a proportion θ of infected individuals will need hospital care, we can then predict, early on in an infection (assuming that the SIR model is correct!), the maximum number $\theta I(t_p)$ of people who will require hospital beds (or intensive care treatment) at any given time, and thus take a more stringent NPI policy if this number is projected to overwhelm hospital capacity.

Let us again take the conservation law

$$I(t) + S(t) - \frac{\nu}{\beta} \ln S(t) \;=\; I_0 + S_0 - \frac{\nu}{\beta} \ln S_0 \,,$$

rearrange and divide by S_0:

$$\frac{I(t)}{S_0} \;=\; \frac{I_0}{S_0} + 1 - \frac{S(t)}{S_0} + \frac{1}{\mathcal{R}_0} \ln\left(\frac{S(t)}{S_0}\right)$$

and now specialize at $t = t_p$, using that $S(t_p) = \frac{\nu}{\beta}$, which means that $\frac{S(t_p)}{S_0} = \frac{\nu}{\beta S_0} = \frac{1}{\mathcal{R}_0}$:

$$\frac{I(t_p)}{S_0} \;=\; \frac{I_0}{S_0} + 1 - \frac{1}{\mathcal{R}_0} + \frac{1}{\mathcal{R}_0} \ln\left(\frac{1}{\mathcal{R}_0}\right) \;=\; \frac{I_0}{S_0} + 1 - \frac{1}{\mathcal{R}_0}\left(1 + \ln \mathcal{R}_0\right) \,.$$

Typically, I_0 is tiny compared to the initial population S_0 (for example there might be just one infective individual in a city of a million inhabitants). Thus we obtain the following elegant formula:

$$\frac{I(t_p)}{S_0} \;\approx\; 1 - \frac{1}{\mathcal{R}_0}\left(1 + \ln \mathcal{R}_0\right)$$

that gives us the peak number of infectives (as a fraction of the initial population) in terms only of $\mathcal{R}_0$. Note how the peak is higher if $\mathcal{R}_0$ is larger, which of course we intuitively expect.

5.0.1 Alternative definitions of $\mathcal{R}_0$

There are alternative definitions of $\mathcal{R}_0$ (for the SIR model) that one often encounters in the literature: $\mathcal{R}_0 = \beta N/\nu$, where N is the total population size, or even $\mathcal{R}_0 = \beta/\nu$. Let us quickly explain how these relate to what we are doing here. As alluded to earlier, the general definition of $\mathcal{R}_0$ is given in terms of what is called a "disease-free steady state" (DFSS), meaning a steady state in which there are no infected individuals. For the specific case of the SIR model, this would mean any steady state of the form $S = S_0$, $I = 0$, and $R = N - S_0$. With this definition, $\mathcal{R}_0 = \beta S_0/\nu$, but there are many possible $\mathcal{R}_0$'s depending on what is the number or removed individuals at the initial time. In particular, for the equilibrium with $R = 0$, $\mathcal{R}_0 = \beta N/\nu$. In the SIRS model, studied later, the only possible DFSS has $R = 0$ (because of the term $-\gamma R$). Thus, for the SIRS model, $\mathcal{R}_0 = \beta N/\nu$.

What about the definition $\mathcal{R}_0 = \beta/\nu$? It is often the case that one normalizes the population to fractions: $S_f := S/N$, $I_f := I/N$, $\widetilde{R} := R/N$. In this case, $\dot{S} = (1/N)(-\beta SI) = -(\beta N)(S_f I_f)$ and $\dot{I}_f = (1/N)(\beta S - \nu)I = ((\beta N)S_f - \nu)I_f$, so

$$\begin{aligned} \dot{S}_f &= -\widetilde{\beta} S_f I_f \\ \dot{I}_f &= (\widetilde{\beta} S_f - \nu) I_f \end{aligned}$$

with $\widetilde{\beta} := \beta N/\nu$. Now $\mathcal{R}_0 = \beta N/\nu = \widetilde{\beta}/\nu$ in terms of this new β. Note that S_f and I_f are dimensionless and that $\widetilde{\beta}$ has units of (1/time), while the original β had units of 1/(time$\times$individuals), so $\widetilde{\beta}$ is perhaps more elegant.

We prefer not to perform this normalization because, when there are "vital dynamics" such as immigration, emigration, births, and/or deaths, the total population N would not be constant.

References

[1] T. P. B. Thu, P. N. Ngoc, N. M. Hai, et al. Effect of the social distancing measures on the spread of COVID-19 in 10 highly infected countries. *Science of the Total Environment*, 742:140430, 2020.

[2] A. L. Wright, K. Sonin, J. Driscoll, and J. Wilson. Poverty and economic dislocation reduce compliance with COVID-19 shelter-in-place protocols. *University of Chicago, Becker Friedman Institute for Economics Working Paper*, (2020-40), 2020.

[3] O. Coibion, Y. Gorodnichenko, and M. Weber. The cost of the COVID-19 crisis: Lockdowns, macroeconomic expectations, and consumer spending. Technical report, National Bureau of Economic Research, 2020.

[4] T. Andersson, A. Erlanson, D. Spiro, and R. Östling. Optimal trade-off between economic activity and health during an epidemic. *arXiv preprint arXiv:2005.07590*, 2020.

[5] W. Kermack and A. McKendrick. A contribution to the mathematical theory of epidemics. *The Royal Society Publishing*, 114(772), 1927.

[6] D. Adam. Special report: The simulations driving the world's response to COVID-19. *Nature*, 580(7803):316–318, 04 2020.

[7] N. Ferguson, D. Laydon, G. Nedjati-Gilani, N. Imai, K. Ainslie, M. Baguelin, S. Bhatia, A. Boonyasiri, Z. Cucunubá, G. Cuomo-Dannenburg, et al. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. *Imperial College, London. DOI: https://doi. org/10.25561/77482*, 2020.

[8] J. Lourenco, R. Paton, M. Ghafari, M. Kraemer, C. Thompson, P. Simmonds, P. Klenerman, and S. Gupta. Fundamental principles of epidemic spread highlight the immediate need for large-scale serological surveys to assess the stage of the SARS-CoV-2 epidemic. *medRxiv*, 2020.

[9] C. Murray. Forecasting the impact of the first wave of the COVID-19 pandemic on hospital demand and deaths for the usa and european economic area countries. *medRxiv*, 2020.

[10] X. Liu, G. Hewings, S. Wang, M. Qin, X. Xiang, S. Zheng, and X. Li. Modeling the situation of COVID-19 and effects of different containment strategies in china with dynamic differential equations and parameters estimation. *medRxiv*, 2020.

[11] M. Chinazzi, J. Davis, M. Ajelli, C. Gioannini, M. Litvinova, S. Merler, A. Pastore y Piontti, K. Mu, L. Rossi, K. Sun, C. Viboud, X. Xiong, H. Yu, M. Halloran, I. Longini, and A. Vespignani. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science*, 368(6489):395–400, 2020.

[12] H. Tian, Y. Liu, Y. Li, C. Wu, B. Chen, M. Kraemer, B. Li, J. Cai, B. Xu, Q. Yang, B. Wang, P. Yang, Y. Cui, Y. Song, P. Zheng, Q. Wang, O. Bjornstad, R. Yang, B. Grenfell, O. Pybus, and C. Dye. The impact of transmission control measures during the first 50 days of the COVID-19 epidemic in china. *medRxiv*, 2020.

[13] A. Kucharski, T. Russell, C. Diamond, Y. Liu, J. Edmunds, S. Funk, R. Eggo, F. Sun, M. Jit, J. Munday, N. Davies, A. Gimma, K. Van Zandvoort, H. Gibbs, J. Hellewell, C. Jarvis, S. Clifford, B. Quilty, N. Bosse, S. Abbott, P. Klepac, and S. Flasche. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect Dis*, 20(5):553–558, 05 2020.

[14] S. Park, D. Cornforth, J. Dushoff, and J. Weitz. The time scale of asymptomatic transmission affects estimates of epidemic potential in the COVID-19 outbreak. *medRxiv*, 2020.

[15] F. Di Lauro, I. Z. Kiss, and J. C. Miller. Optimal timing of one-shot interventions for epidemic control. *PLoS Comput Biol*, 17(3):e1008763, 03 2021.

[16] D.H. Morris, F.W. Rossine, J.B. Plotkin, and S. A. Levin. Optimal, near-optimal, and robust epidemic control. *arXiv*, 2004.02209, 2021.

[17] P. A. Bliman and M. Duprez. How best can finite-time social distancing reduce epidemic final size? *J Theor Biol*, 511:110557, Feb 2021.

[18] E.A. Hernandez-Vargas, G. Giordano, E. D. Sontag, J. G. Chase, H. Chang, and A. Astolfi. First special section on systems and control research efforts against covid-19 and future pandemics. *Annual Reviews in Control*, 50:343–344, 2020.

[19] E.D. Sontag. Notes on mathematical systems biology. `https://drive.google.com/drive/folders/1lIRqaCPeXMVZGoY-44bBsvtnsHtlRfIO?usp=sharing`, 2021.

[20] G. Giordano, F. Blanchini, R. Bruno, P. Colaneri, A. Di Filippo, A. Di Matteo, and M. Colaneri. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nat. Med.*, 26:855860, 2020.

[21] M. Robinson and N. I. Stilianakis. A model for the emergence of drug resistance in the presence of asymptomatic infections. *Mathematical Biosciences*, 243(2):163–177, 2013.

[22] M. Grunnill. An exploration of the role of asymptomatic infections in the epidemiology of dengue viruses through susceptible, asymptomatic, infected and recovered (sair) models. *Journal of Theoretical Biology*, 439:195–204, 2018.

[23] E. Franco. A feedback SIR (fSIR) model highlights advantages and limitations of infection-based social distancing. *arXiv preprint arXiv:2004.13216*, 2020.

[24] M. G. Pedersen and M. Meneghini. Quantifying undetected COVID-19 cases and effects of containment measures in Italy. *ResearchGate Preprint (online 21 March 2020) DOI*, 10, 2020.

[25] J.L. Gevertz, J.M. Greene, C Hixahuary Sanchez Tapia, and E D Sontag. A novel COVID-19 epidemiological model with explicit susceptible and asymptomatic isolation compartments reveals unexpected consequences of timing social distancing. *Journal of Theoretical Biology*, 510:110539, 2020.

[26] M.A. Al-Radhawi, M. Sadeghi, and E.D. Sontag. Long-term regulation of prolonged epidemic outbreaks in large populations via adaptive control: a singular perturbation approach. 2021. Submitted (preprint in http://arxiv.org/abs/2103.08488).

[27] S. Federico and G. Ferrari. Taming the spread of an epidemic by lockdown policies. Center for Mathematical Economics Working Papers 639, Center for Mathematical Economics, Bielefeld University, July 2020.

[28] M. Sadeghi, J.M. Greene, and E.D. Sontag. Universal features of epidemic models under social distancing guidelines. *Annual Reviews in Control*, 2021. To appear (preprint in bioRxiv, 2020, https://www.biorxiv.org/content/10.1101/2020.06.21.163931v2).

[29] E.D. Sontag. *Mathematical Control Theory. Deterministic Finite-Dimensional Systems*, volume 6 of *Texts in Applied Mathematics*. Springer-Verlag, New York, second edition, 1998.

[30] O. Diekmann, J. Heesterbeek, and J. Metz. On the definition and the computation of the basic reproduction ratio r_0 in models for infectious diseases in heterogeneous populations. *J. Math. Biol.*, 28, 1990.

[31] O. Diekmann, J. Heesterbeek, and M. Roberts. The construction of next-generation matrices for compartmental epidemic models. *J. R. Soc. Interface. DOI: 10.1098/rsif.2009.0386*, 7, 2010.